DNA mixtures: biostatistics for mixed stains with haplotypic genetic markers

Fukshansky, N ; Bär, W

Abstract: The conventional theory for interpreting forensic DNA evidence developed for the autosomal genetic markers is not applicable in the case of haplotypic markers, specifically for Y-STR based data. The reason is, that in contrast to the case of autosomal markers, single alleles found in the mixed stain cannot be assigned to unknown stain contributors independently of each other, while the assignable entities are sets of linked alleles which should be treated as non-separable units. It is shown that the conventional theory cannot be extended to this situation. A novel theory which accounts for the features of haplotypic markers has been developed within the general framework of the hypotheses testing approach. This theory opens the way for the use of haplotypic markers in the analysis of mixed stains with the arbitrary numbers of unknown contributors and linked loci. A numerical example demonstrates the application of the theory

DOI: https://doi.org/10.1007/s00414-004-0497-5

Int J Legal Med (2005) 119: 285–290
DOI 10.1007/s00414-004-0497-5

TECHNICAL NOTE

N. Fukshansky · W. Bär

DNA mixtures: biostatistics for mixed stains with haplotypic genetic markers

Received: 22 March 2004 / Accepted: 27 September 2004 / Published online: 17 February 2005

Abstract The conventional theory for interpreting forensic DNA evidence developed for the autosomal genetic markers is not applicable in the case of haplotypic markers, specifically for Y-STR based data. The reason is, that in contrast to the case of autosomal markers, single alleles found in the mixed stain cannot be assigned to unknown stain contributors independently of each other, while the assignable entities are sets of linked alleles which should be treated as non-separable units. It is shown that the conventional theory cannot be extended to this situation. A novel theory which accounts for the features of haplotypic markers has been developed within the general framework of the hypotheses testing approach. This theory opens the way for the use of haplotypic markers in the analysis of mixed stains with the arbitrary numbers of unknown contributors and linked loci. A numerical example demonstrates the application of the theory.

Keywords DNA · Forensic statistics · Mixed stains · Y-STR haplotypes

Introduction

Rapid accumulation of data concerning polymorphisms of human Y-chromosomes opens the way for applications of Y-chromosome profiling and Y-STR-based haplotype distributions for purposes of molecular phylogenetics, analysis of geographic population structures, personal identification as well as various kinship and forensic studies (Jobling et al. 1997; Forster et al. 1998; Pestoni et al. 1999; Roewer et al. 2000; Kayser et al. 2002; Gusmao et al. 2003). The features of Y-markers, principally haploidy and the absence of recombination, provide a number of advantages but also limitations both being problem dependent. Also, in forensic stain analysis the application of Y-chromosomal microsatellites is widely used in routine work (Sparkes et al. 1996a,b; Redd et al. 1997; Pascali et al. 1998; Schneider et al. 1999; Roewer and Carracedo 2001; Gill et al. 2001; Zarrabeitia et al. 2003). However, the analysis of mixed stains in the case of non-exclusion cannot yet be dealt with. Here the following field-specific theoretical problem arises which, to our knowledge, has not yet been addressed.

The purpose of conventional stain analysis is to provide evidence concerning the crime scenario, where we compare the mixed genetic material from the crime scene (stain) to the single person tests of stain contributors (victims, suspected assailants). In the case of non-exclusion the analysis appears rather complicated, since the existence of non-tested persons (unknowns) contributing to the stain has to be taken into consideration. For the autosomal genetic markers there exists a well established theory which is derived within the general scope of the hypotheses testing approach (Weir et al. 1997; Fukshansky and Bär 1998).

In this paper we will show that if a mixed stain comprises haplotypic (specifically Y-STR based) markers, the arguments leading to the conventional theory as developed for the autosomal markers, in principle, cannot be used. The consequence is that the conventional theory can be neither directly applied nor modified for the application to these stains. We also propose a corresponding feasible solution for haplotype systems in general.

The analysis of haplotype-based results from stains is formulated in general terms of the hypotheses testing approach. The general theory for haplotype-based stains is developed starting from the elementary case "one unknown person and two-locus haplotype" by means of recursions with respect to the number of unknowns and then these recursions are extended to the case of an arbitrary number of

N. Fukshansky
Birkenstrasse 4B,
79256 Buchenbach, Germany
e-mail: lefu@ruf.uni-freiburg.de

W. Bär (✉)
Institut für Rechtsmedizin,
Universität Zürich-Irchel,
Winterthurerstrasse 190/Bau 52,
8057 Zürich, Switzerland
e-mail: baer@irm.unizh.ch
Tel.: 0041-1-6355621
Fax: 0041-1-6356815

linked loci. A numerical example of application is given together with some concluding remarks and a future outlook.

Analysis of Y-chromosome-based mixed stains as a problem of hypotheses testing

Irrespective of the nature of genetic markers, each hypothesis in mixture evaluation is a statement specifying members of a group of persons—among all tested persons and, if necessary, non-tested persons (unknowns), as either contributors or non-contributors to the stain. The statistical analysis has the two following limitations: all unknown persons should belong to the same ethnic group and there should be no genetical relationships between them.

In the general case of a hypotheses testing concerning a crime scenario, the calculation of the probability for the genetic profile of a stain S if it contains DNA from n unknown contributors and some number of tested contributors with personal tests $T_1,T_2,...$ reduced to the computation of the conditional probability that the n unknowns are contributors to the stain S under the condition that tested persons with personal tests $T_1,T_2,...$ are also contributors to the stain (Fukshansky and Bär 1998). We will specify this conditional probability as

$$P_n(S|T_1, T_2, ...), \qquad \text{where} \qquad P_0(S|T_1, T_2, ...) = 1$$

In the case of autosomal markers (non-linked loci), when two alleles are assigned to each person, a rather simple formula (Weir's formula) has been derived for the computation of the probability $P_n(S|T_1,T_2,...)$ (Weir et al. 1997; Fukshansky and Bär 1998). This derivation is essentially based on the fact that each allele found in the stain can be assigned to any of the unknown contributors independently of other alleles. This independence disappears as soon as we are dealing with the Y-chromosomal markers. Here the stain is described, as in the autosomal case, by a set of separate alleles, however, the assignable entities, are haplotypes, i.e. sets of linked alleles, and must be handled as non-separable units.

In order to formulate the hypotheses testing formalism in this more complicated situation we proceed with a number of simple detailed case studies. Let us start with a haplotype containing three loci A,B,C with the sets of alleles

$$A_1, A_2, ..., A_{n_1} \quad B_1, B_2, ..., B_{n_2} \quad C_1, C_2, ..., C_{n_3}$$

and therefore $n_1\ n_2\ n_3$ different haplotypes. We designate a haplotype $A_iB_jC_k$ simply as (ijk), its frequency in the population as $f(ijk)$ and the probability that n unknowns are the contributors to the stain S as $p_n(S)$.

Further, we introduce a similar simplification for the description of a stain. For example the stain S=A_2A_3,B_1B_2, $C_1C_2C_3$ will be specified as (23,12,123) and the probability that n unknowns are contributors to this stain is now

$$p_n(S) = p_n(23, 12, 123).$$

Let us further introduce the following useful definitions:

1. A locus in a stain with N contributors is called *complete* if the stain contains N alleles of this locus, otherwise the locus is called *incomplete*
2. A stain is called *complete* if all the loci contained in the stain are complete, otherwise the stain is called *incomplete*
3. A stain is called *absolutely incomplete* if all the loci contained in the stain are incomplete.

Now we can consider examples of complete and incomplete stains for the case of three contributors (N=3) among which two are unknowns (n=2).

1. Let us consider the complete stain S=(123,123,123). The haplotype of the only tested person is T=(111). In this situation only one possibility remains for the two unknowns: they must show together the complete stain (23,23,23). The probability of this event is designated as p_2(23,23,23). Thus, the probability that the two unknowns are contributors to the stain S under the condition that one tested person with the personal test result T also contributes to the stain is

$$P_3(123, 123, 123|(111)) = p_2(23, 23, 23).$$

2. Now let us consider the incomplete stain S=(12, 123,123). Again, the haplotype of the only tested person is T=(111). Now the two unknowns must show together the alleles A_2,B_2,B_3,C_2C_3 and they can show in addition only alleles from the stain. This means that they can show together the incomplete stain (2,23,23) or the complete stain (12,23,23), so that the corresponding conditional probability is

$$P_3(12, 123, 123|(111)) = p_2(2, 23, 23) + p_2(12, 23, 23).$$

These examples show that the sought conditional probability turns out to be a sum of non-conditional probabilities that n unknowns are contributors to various stains which are reductions of the original stain (in some cases, specifically when the original stain is a complete one, this sum reduces to a single item). The nature and the number of these items are determined by the correlations between the original stain and the personal results of the tested persons. It is important to emphasize that all the items are absolute (non-conditional) probabilities, $p_n(S)$, (the probabilities that the n unknowns are contributors to various stains S).

In the next two sections we will construct $p_n(S)$ for all possible forms of S on the basis of formulas which are recursive with respect to the number of unknowns. First, this treatment will be done for the two-loci haplotype and then extended to the arbitrary number of linked loci.

Stain probability—recursion for the number of unknown contributors in a two-loci haplotype

We start the derivation with the simplest haplotype consisting of two loci, *A* and *B*.

The basis for the recursion is n=1, i.e. only one unknown contributor to the stain. A single person can contribute only his own haplotype, i.e. a complete stain $A_i,B_j=(i,j)$, which appears with the probability $p_1(i,j)=f(ij)$, where $f(ij)$ is the frequency of the haplotype (ij) in the population as specified previously.

For two unknowns (n=2) one should consider the complete stain $A_{i1}A_{i2},B_{j1}B_{j2}=(i_1i_2,j_1j_2)$ as well as the incomplete stains for the cases when the unknowns have identical alleles in one locus, $A_{i1},B_{j1}B_{j2}=(i_1,j_1j_2)$ and $A_{i1}A_{i2},B_{j1}=(i_1i_2,j_1)$ or in both loci, $A_{i1},B_{j1}=(i_1,j_1)$. Before deriving the probabilities for these stains we will simplify some more specifications: we will indicate the stain $A_{i1}A_{i2},B_{j1}B_{j2}$ as (12,12) and, correspondingly, the stains $A_{i1},B_{j1}B_{j2}$, $A_{i1}A_{i2}$, B_{j1}, A_{i1},B_{j1} as (1,12), (12,1), (1,1) respectively.

Let us consider the stain S=(12,12). Four haplotypes may be involved in this stain:

$$(11), (12), (21), (22).$$

For each fixed haplotype of the first unknown there exists only one haplotype of the second unknown, which supplements it to the original stain *S*. This yields the probability for *S* with two unknowns

$$p_2(12,12) = f(11)p_1(2,2) + f(12)p_1(2,1) + f(21)p_1(1,2) + f(22)p_1(1,1).$$

Note that the supplementary stain of the second unknown arises from the original stain *S* by means of canceling alleles shown by the first unknown. For example, the supplementary stain (2,2) arises by canceling alleles A_1 and B_1 in the stain (12,12). Let us generally specify the supplementary stain generated by canceling the alleles A_i, B_j as S_{ij}. Applying this specification for each supplementary stain we can rewrite the last formula as

$$p_2(S=12,12) = \sum_{i=1}^{2}\sum_{j=1}^{2} f(ij)p_1(S_{i,j}).$$

Let us now consider the stain S=(1,12). Two haplotypes may be involved in this stain: (11) and (12). Therefore

$$p_2(1,12) = f(11)p_1(1,2) + f(12)p_1(1,1).$$

Now we see that the supplementary stain shown by the second unknown is generated by canceling alleles in the second locus alone, whereas no alleles were canceled in the first locus (an obvious consequence of the incompleteness of the original stain). To account for such cases let us extend our specification of the supplementary stain by introducing terms like $p_1(S_{i,0})$ and $p_1(S_{0,j})$ where a zero at a certain locus position means that no canceling was performed on this locus. With this extension we can rewrite the last formula as

$$p_2(S=1,12) = \sum_{i=1}^{1}\sum_{j=1}^{2} f(ij)p_1(S_{0,j}).$$

Let the stain *S* be S=(1,1). A single haplotype (11) covers this stain. The two unknowns possess identical haplotypes and the expression for the probability of the original stain reads

$$p_2(S=1,1) = \sum_{i=1}^{1}\sum_{j=1}^{1} f(ij)p_1(S_{0,0}).$$

Before we derive the general expressions, let us also consider the case n=3. For the complete stain S=(123,123) reasoning as in the above examples yields

$$p_3(S=123,123) = \sum_{i=1}^{3}\sum_{j=1}^{3} f(ij)p_2(S_{i,j}).$$

Let us now consider the incomplete stain S=(12,123). If the first unknown has the haplotype (11), then the two remaining unknowns must provide together the alleles A_2, B_2,B_3 and in addition they can have other alleles from the stain *S*. This means that together they can be the contributors to the stain (12,23) with corresponding probability p_2 (S=12,23) $p_2(S_{0,1})$ or to the stain (2,23) with the probability $p_2(2,23)=p_2(S_{1,1})$. Applying the same arguments for all the possible fixations of the first unknowns' haplotype we get

$$p_3(S=12,123) = \sum_{i=1}^{2}\sum_{j=1}^{3} f(ij)[p_2(S_{0,j}) + p_2(S_{i,j})].$$

Let us consider now the absolutely incomplete stain S=(12,12). If the first unknown has the haplotype (11) then the two remaining unknowns must provide together the alleles A_2 and B_2 and in addition they can have other alleles from the stain *S*. This means that together they can be the contributors to the stains (12,12), (2,12), (12,2) and (2,2). This yields the formula

$$p_3(S=12,12) = \sum_{i=1}^{2}\sum_{j=1}^{2} f(ij)[p_2(S_{0,0}) + p_2(S_{i,0}) + p_2(S_{0,j}) + p_2(S_{i,j})].$$

The following expressions are now obvious

$$p_3(S=1,123) = \sum_{i=1}^{1}\sum_{j=1}^{3} f(ij)p_2(S_{0,j})$$

$$p_3(S=1,12)=\sum_{i=1}^{1}\sum_{j=1}^{2}f(ij)[p_2(S_{0,0})+p_2(S_{0,j})]$$

$$p_3(S=1,1)=\sum_{i=1}^{1}\sum_{j=1}^{1}f(ij)p_2(S_{0,0})$$

On the basis of the above examples we proceed now to the general derivation of the transition from $n-1$ to n unknowns for $n\geq 3$, but still restricted to the case of two linked loci.

The set of all possible stains for n unknowns can be described as

$$(1\,2...m_1,\,1\,2...m_2)\quad\text{with}\quad 1\leqslant m_1,m_2\leqslant n.$$

If a locus in the original stain S is complete then after the haplotype of one unknown becomes fixed, this locus also appears complete in the supplementary stain for the rest of the $n-1$ unknowns. This means that the corresponding coordinate in the item $p_{n-1}(S_{i,j})$ cannot be zero. If a locus is incomplete and the number of its alleles is larger than 1, two possibilities—zero and non-zero—should be taken into consideration for this coordinate. Finally, if an incomplete locus is presented in S by a single allele, the corresponding coordinate must be zero. Thus, the transition from $n-1$ to n unknowns will be given by the following expressions. In these formulas we assume that $1<m_1,m_2<n$.

When at least one locus is complete one has:

$$p_n(S=12...n,12...n)=\sum_{i=1}^{n}\sum_{j=1}^{n}f(ij)p_{n-1}(S_{i,j})$$

$$p_n(S=12...m_1,12...n)=\sum_{i=1}^{m_1}\sum_{j=1}^{n}f(ij)[p_{n-1}(S_{0,j})+p_{n-1}(S_{i,j})]$$

$$p_n(S=1,12...n)=\sum_{i=1}^{1}\sum_{j=1}^{n}f(ij)p_{n-1}(S_{0,j})\qquad(1)$$

and when the stain is absolutely incomplete:

$$p_n(S=12...m_1,12...m_2)=\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}f(ij)[p_{n-1}(S_{0,0})+p_{n-1}(S_{i,0})+p_{n-1}(S_{0,j})+p_{n-1}(S_{i,j})]$$

$$p_n(S=1,12...m_2)=\sum_{i=1}^{1}\sum_{j=1}^{m_2}f(ij)[p_{n-1}(S_{0,0})+p_{n-1}(S_{0,j})]$$

$$p_n(S=1,1)=\sum_{i=1}^{1}\sum_{j=1}^{1}f(ij)p_{n-1}(S_{0,0})\qquad(2)$$

These expressions give the complete solution for haplotypes consisting of two loci. In the next section we proceed to the last step—extension to the arbitrary number of linked loci.

Stain probability—extension to the arbitrary number of linked loci

Before arriving at a general formula for m linked loci we will demonstrate the transition from $m-1$ to m for the case $m=3$, where we take the formulas (1, 2): for the case of three linked loci the set of all possible stains for n unknown contributors can be described as

$$(1\,2...m_1,\,1\,2...m_2,\,1\,2...m_3)\quad\text{with}\quad 1\leqslant m_1,m_2,m_3\leqslant n$$

For the transition from $m=2$ to $m=3$ the haplotype frequencies $f(ijk)$ $(i,j,k>0=)$ and the supplementary stains, S_{ijk}, acquire the third coordinate.

If the additional (third) locus is complete then, as discussed in the previous section, the term $p_{n-1}(S_{i,j,k})$ has k>0. Therefore, a natural extension of the formulas (1) yields the corresponding expression for the three-loci stains having at least two complete loci. In the following formulas we assume that $1<m_1,m_2,m_3<n$.

$$p_n(S=1...n,1...n,1...n)=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}f(ijk)p_{n-1}(S_{i,j,k})$$

$$p_n(S=1...m_1,1...n,1...n)=\sum_{i=1}^{m_1}\sum_{j=1}^{n}\sum_{k=1}^{n}f(ijk)\times[p_{n-1}(S_{0,j,k})+p_{n-1}(S_{i,j,k})]$$

$$p_n(S=1,1...n,1...n)=\sum_{i=1}^{1}\sum_{j=1}^{n}\sum_{k=1}^{n}f(ijk)p_{n-1}(S_{0,j,k})\qquad(3)$$

In the same way the formulas (2) extend to the three-loci stains having one complete locus:

$$p_n(S=1...m_1,1...m_2,1...n)=\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{k=1}^{n}f(ijk)\times[p_{n-1}(S_{0,0,k})+p_{n-1}(S_{i,0,k})+p_{n-1}(S_{0,j,k})+p_{n-1}(S_{i,j,k})]$$

$$p_n(S=1,1...m_2,1...n)=\sum_{i=1}^{1}\sum_{j=1}^{m_2}\sum_{k=1}^{n}f(ijk)\times[p_{n-1}(S_{0,0,k})+p_{n-1}(S_{0,j,k})]$$

$$p_n(S=1,1,1...n)=\sum_{i=1}^{1}\sum_{j=1}^{1}\sum_{k=1}^{n}f(ijk)p_{n-1}(S_{0,0,k})\qquad(4)$$

Finally, for an absolutely incomplete stain the formulas (4) convert to expressions containing in addition to items $p_{n-1}(S_{i,j,k})$, terms of the type $p_{n-1}(S_{i,j,0})$:

$$p_n(S=1...m_1,1...m_2,1...m_3)=\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{k=1}^{m_3}f(ijk)\times[p_{n-1}(S_{0,0,0})+p_{n-1}(S_{0,0,k})+p_{n-1}(S_{i,0,0})+p_{n-1}(S_{i,0,k})+p_{n-1}(S_{0,j,0})+p_{n-1}(S_{0,j,k})+p_{n-1}(S_{i,j,0})+p_{n-1}(S_{i,j,k})]$$

$$p_n(S=1,1...m_2,1...m_3)=\sum_{i=1}^{1}\sum_{j=1}^{m_2}\sum_{k=1}^{m_3}f(ijk)\times[p_{n-1}(S_{0,0,0})+p_{n-1}(S_{0,0,k})+p_{n-1}(S_{0,j,0})+p_{n-1}(S_{0,j,k})]$$

$$p_n(S=1,1,1...m_3)=\sum_{i=1}^{1}\sum_{j=1}^{1}\sum_{k=1}^{m_3}f(ijk)[p_{n-1}(S_{0,0,0})+p_{n-1}(S_{0,0,k})]$$

$$p_n(S=1,1,1)=f(111)p_{n-1}(S_{0,0,0}) \quad (5)$$

Applying the same reasoning we can now write the overall formula for M linked loci:

$$p_n(S=1...m_1,...,1...m_j,...,1...m_M)=\sum_{i_1=1}^{m_1}...\sum_{i_j=1}^{m_j}...\sum_{i_M=1}^{m_M}f(i_1,...,i_j,...,i_M)\times\sum p_{n-1}(S_{i_1,...,i_j,...i_M}), \quad (6)$$

where the variables $i_j(j=1,...M)$ in the expression $p_{n-1}(S_{i1,...,jj,...im})$ acquire the following values:

$$i_j \quad \begin{matrix} i_j & \text{for} & m_j=n \\ \text{and} \quad 0 & \text{for} & 1<m_j<n \\ 0 & \text{for} & m_j=1 \end{matrix}$$

Numerical example

As an example of the application of the theory let us consider the case of three linked loci and the stain

$$S=A_1,B_1B_2,C_1C_2C_3=(1,12,123)$$

produced by $N=3$ contributors. The two non-excluded suspects, S_1 and S_2, have been tested and showed the following haplotypes

$$T_1=(111) \qquad \text{and} \qquad T_2=(112).$$

In order to analyse the crime scenarios we have to consider the following four hypotheses:

H_1 The stain is produced by the two suspects and one unknown

$$=P_1(1,12,123|T_1,T_2)=p_1(1,2,3)=f(123)$$

H_2 S_1 is not the contributor to the stain, which is produced by S_2 and two unknowns

$$=P_2(1,12,123|T_2)=p_2(1,12,13)+p_2(1,2,13)$$

H_3 S_2 is not a contributor to the stain, which is produced by S_1 and two unknowns.

$$=P_2(1,12,123|T_1)=p_2(1,12,23)+p_2(1,2,23)$$

H_4 S_1 and S_2 are not contributors to the stain, which is produced by three unknowns

$$=p_3(1,12,123).$$

Using the second formula of (4) for $n=3$ and moving on from $p_2(S_{ij,k})$ to the corresponding values p_2 we obtain

$$\begin{aligned}p_3(1,12,123)&=\sum_{j=1}^{2}\sum_{k=1}^{3}f(1jk)[p_2(S_{0,0,k})+p_2(S_{0,j,k})]\\&=f(111)[p_2(1,12,23)+p_2(1,2,23)]\\&\quad+f(112)[p_2(1,12,13)+p_2(1,2,13)]\\&\quad+f(113)[p_2(1,12,12)+p_2(1,2,12)]\\&\quad+f(121)[p_2(1,12,23)+p_2(1,1,23)]\\&\quad+f(122)[p_2(1,12,13)+p_2(1,1,13)]\\&\quad+f(123)[p_2(1,12,12)+p_2(1,1,12)].\end{aligned}$$

The probabilities p_2 can be computed according to the formulas:

$$p_2(1,12,k_1k_2)=2\,[f(1\,1\,k_1)f(1\,2\,k_2)+f(1\,1\,k_2)f(1\,2\,k_1)]$$

$$p_2(1,j,k_1k_2)=2f(1\,j\,k_1)f(1\,j\,k_2)$$

Let us assume the following tabulated frequencies in the corresponding population:

$$f(111)=2.88\ 10^{-4} \qquad f(112)=0.13\ 10^{-4} \qquad f(113)=0.22\ 10^{-4}$$

$$f(121)=3.60\ 10^{-4} \qquad f(122)=4.32\ 10^{-4} \qquad f(123)=5.04\ 10^{-4}$$

This yields for the probabilities of the hypotheses

$$X_1 = 5.04 \cdot 10^{-4} \qquad X_2 = 8.09 \cdot 10^{-7}$$
$$X_3 = 7.67 \cdot 10^{-7} \qquad X_4 = 1.09 \cdot 10^{-9}$$

and for the likelihood ratios $L_i = X_1/X_i$ for the hypothesis H_i as compared to the hypothesis H_1;

$$L_2 = 6.23 \cdot 10^{2} \qquad L_3 = 8.56 \cdot 10^{2} \qquad L_4 = 4.63 \cdot 10^{5}.$$

Concluding remarks and outlook

The content of this report can be summarized as the two following messages.

1. The existing theory of forensic mixed stain analysis developed for the autosomal genetic markers can be neither directly applied nor modified for application to haplotypic markers (specifically to Y-STR based data).
2. A novel theory which accounts for the features of haplotypic markers has been developed within the general framework of the hypotheses testing approach.

After implementation of this theory as a software package the utilization of the advantages introduced by Y-STR based data will be possible by a sound computational procedure.

However, this utilization, or more exactly the discriminative power of the stain analysis is also dependent on the structure of the haplotype distributions in the population. Here the following theoretical problem arises. Given the typical tabulated haplotype frequencies, what number of linked loci will be sufficient to achieve some aimed stage of discrmination between the hypotheses? This so-called sensitivity analysis will provide the practical basis for the theory and specify its validity range. Finally, the two following problems, which have been already solved for the autosomal markers (Fukshansky and Bär 1999, 2000) should be approached: what modifications of the theory are necessary when the assumed stain contributors are genetically related or belong to a different ethnic group?

References

Cali F, Forster P, Kersting C, Mirisola MG, D'Anna R, De Leo G, Romano V (2002) DXYS156: a multi-purpose short tandem repeat locus for determination of sex, paternal and maternal geographic origins and DNA fingerprinting. Int J Legal Med 116:133–138

Forster P, Kayser M, Meyer E, Roewer L, Pfeiffer H, Benkmann H, Brinkmann B (1998) Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans. Mol Biol Evol 15:1108–1114

Fukshansky N, Bär W (1998) Interpreting forensic DNA evidence on the basis of hypotheses testing. Int J Legal Med 111:62–66

Fukshansky N, Bär W (1999) Biostatistical evaluation of mixed stains with contributors of different ethnic origin. Int J Legal Med 112:383–387

Fukshansky N, Bär W (2000) Biostatistics for mixed stains: the case of tested relatives of a non-tested suspect. Int J Legal Med 114:78–82

Gill P, Brenner C, Brinkmann B et al. (2001) DNA commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. Int J Legal Med 114:305–309

Gusmao L, Sánchez-Diz P, Alves C, Beleza S, Lopes A, Carracedo A, Amorim A (2003) Grouping of Y-STR haplotypes discloses European geographic clines. Forensic Sci Int 134:172–179

Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124

Kayser M, Brauer S, Willuweit S et al. (2002) Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U.S. populations. J Forensic Sci 47:513–519

Pascali VL, Dobosz M, Brinkmann B (1998) Coordinating Y-chromosomal STR research for the courts. Int J Legal Med 112:1

Pestoni C, Cal ML, Lareu MV, Rodriguez-Calvo MS, Carracedo A (1999) Y chromosome STR haplotypes: genetic and sequencing data of the Galician population (NW Spain). Int J Legal Med 112:15–21

Redd AJ, Clifford SL, Stoneking M (1997) Multiplex DNA typing of short-tandem-repeat loci on the Y chromosome. Biol Chem 378:923–927

Roewer L, Carracedo A (2001) Second Forensic Y-chromosome User Workshop. Forensic Sci Int 118:105–181

Roewer L, Kayser M, de Knijff P et al. (2000) A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. Forensic Sci Int 114:31–43

Roewer L, Krawczak M, Willuweit S et al. (2001) Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci Int 118:106–113

Rosser ZH, Zerjal T, Hurles ME et al. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526–1543

Schneider PM, d'Aloja E, Dupuy BM et al. (1999) Results of collaborative study regarding the standardization of the Y-linked STR system DYS385 by the European DNA Profiling (EDNAP) group. Forensic Sci Int 102:159–156

Sparkes R, Kimpton C, Gilbard S et al. (1996b) The validation of a 7-locus multiplex STR test for use in forensic casework. (II), Artefacts, casework studies and success rates. Int J Legal Med 109:195–204

Sparkes R, Kimpton C, Watson S et al. (1996a) The validation of a 7-locus multiplex STR test for use in forensic casework. (I). Mixtures, ageing, degradation and species studies. Int J Legal Med 109:186–94

Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KA, Buckleton J (1997) Interpreting DNA mixtures. J Forensic Sci 42:213–219

Zarrabeitia MT, Riancho JA, Gusmao L, Lareu MV, Sanudo C, Amorim A, Carracedo A (2003) Spanish population data and forensic usefulness of a novel Y-STR set (DYS437, DYS438, DYS439, DYS460, DYS461, GATA A10, GATA C4, GATA H4). Int J Legal Med 117:306–311